



UNIVERSIDAD
PANAMERICANA

Introducción a la Ciencia de Datos

Dr. Leon Felipe Palafox Novack
lpalafox@up.edu.mx

1

Anuncios parroquiales

Por que todos queremos saber como
se va a calificar

“ *A mi no me importan las buenas calificaciones, mis creaciones seran el testamento de mi desempeño.*

Helmut Kohl



Objetivo

El curso inducirá al alumno a la ciencia de datos y proporcionará conocimientos y habilidades para utilizar las diferentes herramientas de Inteligencia de Negocios para generar valor y dar soporte a la toma de decisiones.

Se detallarán mejores prácticas y se dará una introducción al ciclo de vida de un proyecto de Ciencia de Datos.

Requisitos



- Fundamentos de programación, manejo avanzado de Excel, conocimientos básicos de estadística.

Población Objetivo

Directivos o personas de nivel gerencial, analistas que quieran introducirse a la ciencia de datos y en particular obtener habilidades para implementar estrategias de Inteligencia de Negocio.



Calificación

La evaluación consistirá en:

■ El proyecto final será el 60% de la evaluación final.

- ▷ El proyecto final consistirá en el uso de un set de datos de su preferencia para diseñar un caso de negocio.
- ▷ Pueden hacer equipos de hasta tres personas.
- ▷ Necesitan hacer un reporte de 3-5 paginas sobre el set de datos, el diseño y las variables usadas.

■ El restante 40% será distribuido de la siguiente forma:

- ▷ Dos exámenes.
- ▷ Dos tareas.
- ▷ Participación en clase
- ▷ Score en Datacamp

■ https://leonpalafox.github.io/dsclase_2019_2/

- Es una plataforma online para aprender a escribir código en Python.
 - ▷ Es necesario para esta clase.
 - ▷ El link esta en Moodle

2

Introducción

Qué es y con que se come la Data Science?

“ The goal is to turn data into information, and information into insight

Carly Fiorina

Historia de la Ciencia de datos



■ Qué es un Data Scientist?

- ▷ No es un analista
- ▷ No es un programador
- ▷ No es un programador de base de datos.

■ Hace Ciencia?

- ▷ No nos importa

Historia de la Ciencia de datos



DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

 SUMMARY  SAVE  SHARE  COMMENT  TEXT SIZE  PRINT **\$8.95** BUY COPIES



When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing

TECHNOLOGY

These data scientists are disrupting Disney World's long wait times

Touring Plans goes to impressive lengths to help you avoid ride lines.

By Carlye Wisel October 31, 2017



■ La historia de la Ciencia de Datos tiene dos ramas:

- ▷ Machine Learning – Inteligencia Artificial
- ▷ Infraestructura Computo Paralelo

Machine Learning

- ▷ Es el nombre comercial de una rama de Inteligencia Artificial.
- ▷ En 1984 aconteció el primer invierno de la Inteligencia Artificial.
 - ▷ Se prometieron muchas cosas que no se entregaron
 - ▷ Faltaban dos cosas:
 - ▷ Poder de procesamiento
 - ▷ Datos!!!
- ▷ Que creen ustedes que pasó después??

Machine Learning

- En los 2000, todo mundo contrataba desarrolladores de bases de datos.

- ▶ La aparición del developer Full Stack
- ▶ Una acumulación masiva de datos

La aparición de las redes sociales masivas.

- ▶ Facebook tiene alrededor de 2 mil millones de usuarios
- ▶ Se generan datos iguales a alrededor de 500, 000, 000 millones de canciones cada día.
- ▶ Quien los va categorizar?

De pronto a hay datos de sobra, y adivinen que más, el procesamiento también es sustancialmente mayor al que se tenía en 1984.

■ Popularidad de “nuevos” algoritmos:

- ▷ Sistemas de recomendación
- ▷ Redes Neuronales Profundas (Deep Learning)
- ▷ Gradiente Descendente (XGBoost)

- Se abandonó la dependencia en maquinas únicas poderosas.
- El computo distribuido se volvió barato y razonable.
 - ▷ Hadoop
 - ▷ Hortonworks
 - ▷ Cloudera
 - ▷ AWS
 - ▷ Google
 - ▷ Azure

■ Almacenar muchos datos es difícil:

- ▷ Bases de datos paralelas
- ▷ Almacenamiento redundante
- ▷ Almacenamiento distribuido.

With less computing power than a washing machine

! This video is no longer available.

In the mid-60s, a golden generation of highly trained whizz kids was pouring out of American universities with PhDs in maths, engineering and chemistry. "It was the generation that went on to drive the development of silicon valley in the 1970s," says Dave Parker, director of the British National [Space](#) Centre. "And those people made the moon landing happen."

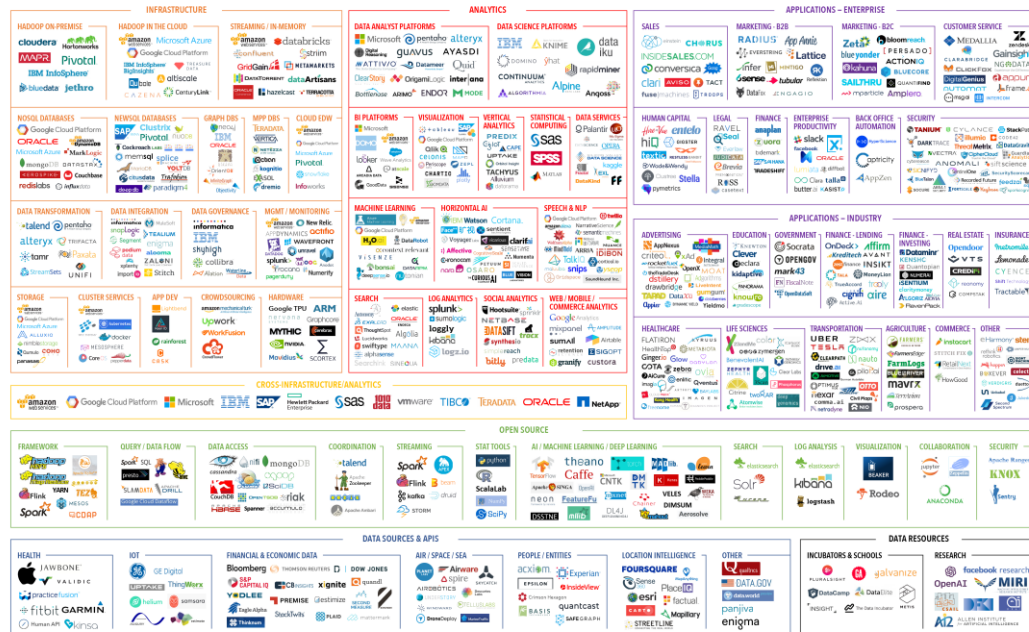
Wernher von Braun, spirited from postwar Germany in Operation Paperclip

Conclusión

- Se requirieron muchas condiciones para la aparición de Data Science como profesión:
 - ▷ Expertise en Machine Learning
 - ▷ Expertise en Hardware
 - ▷ **Expertise en Negocios**

Ecosistema de la Ciencia de Datos

BIG DATA LANDSCAPE 2017

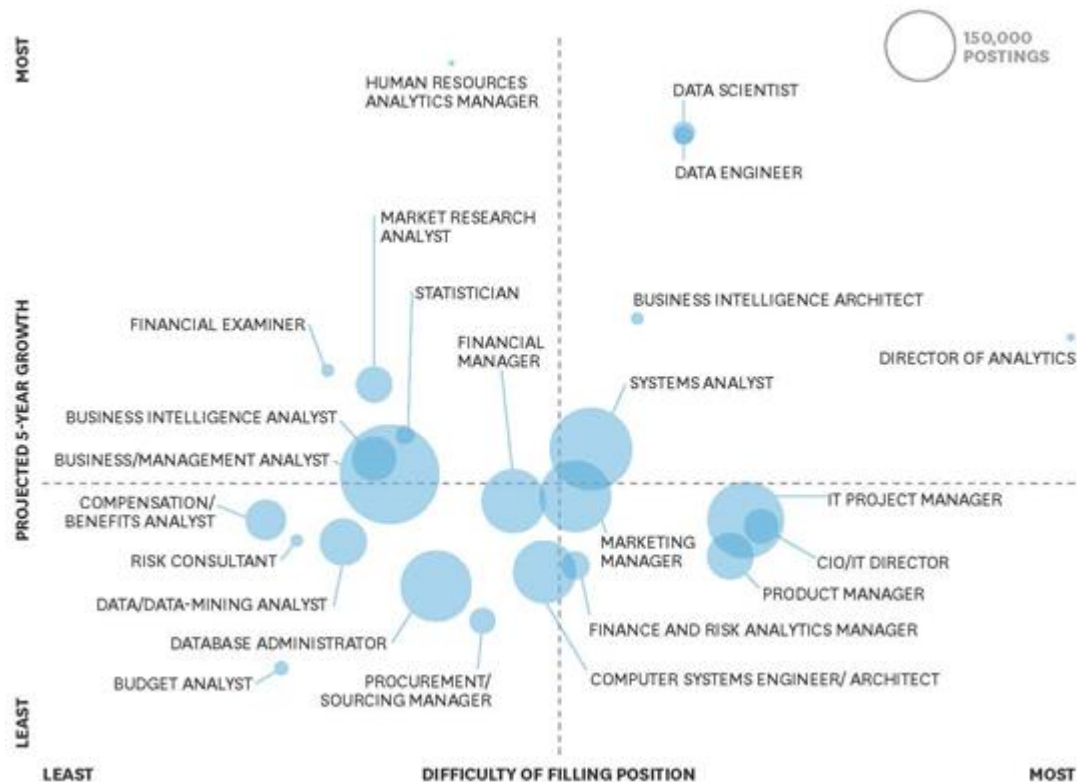


V2 - Last updated 5/3/2017

© Matt Turck (@mattturck), Jim Hao (@jimhao), & FirstMark (@firstmarkcap) mattturck.com/bigdata2017

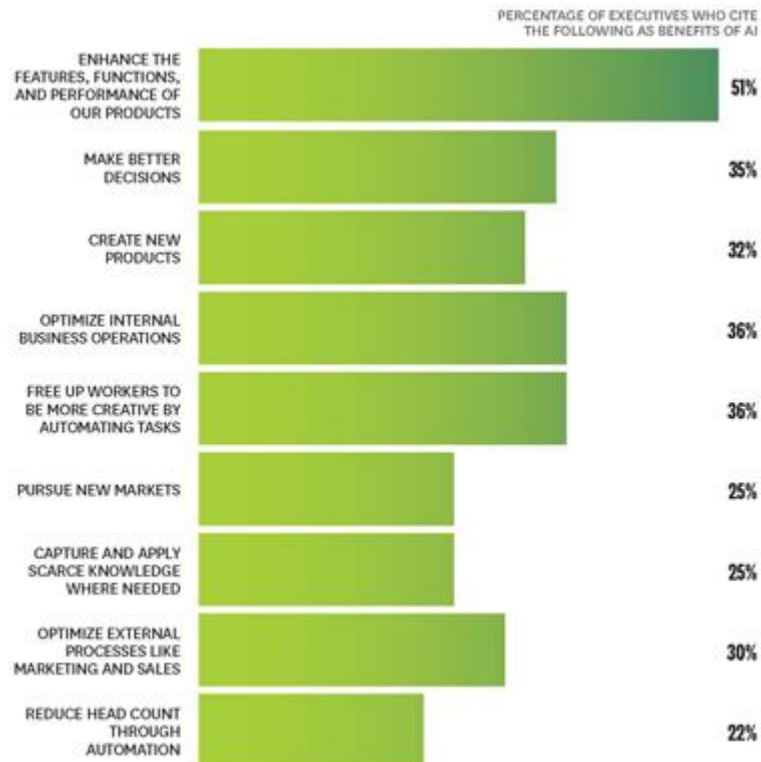
FIRSTMARK
EARLY STAGE VENTURE CAPITAL

<http://mattturck.com/wp-content/uploads/2017/05/Matt-Turck-FirstMark-2017-Big-Data-Landscape.png>



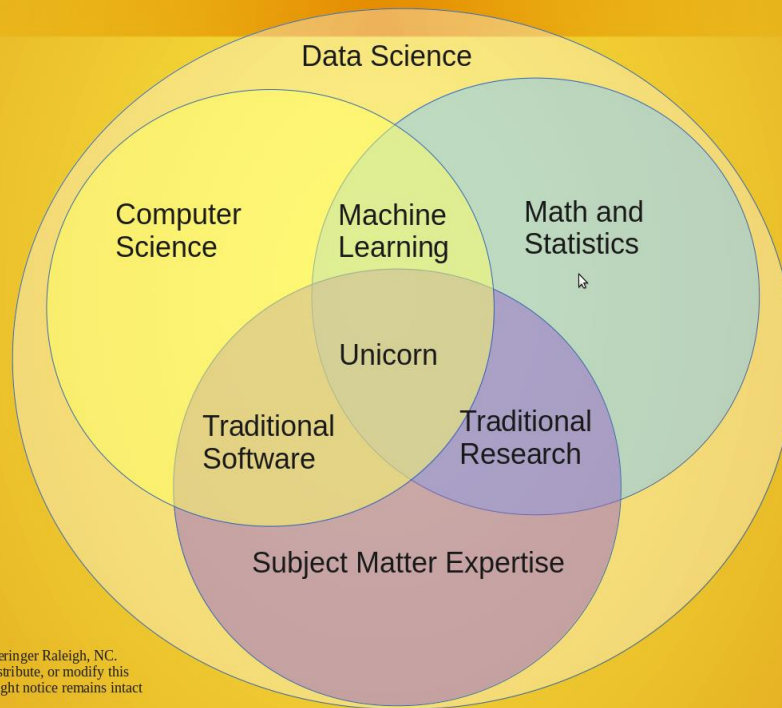


UNIVERSIDAD
PANAMERICANA



SOURCE: DELOITTE 2017

Data Science Venn Diagram v2.0



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this
image, provided that this copyright notice remains intact

3

Casos de uso de la Ciencia de Datos

Donde se come la Data Science?

“ The world is now awash in data
and we can see consumers in a
lot clearer ways.

Max Levchin, PayPal co-founder

Ciencia de Datos en el Mercado de Servicios - Telecomunicaciones



Notre Dame announces collaboration with AT&T for online master's degree in data science

by Sue Lister

August 30, 2016

In a data-driven economy, industry leaders rely increasingly on skilled professionals who can see the significance in data and use it to solve business challenges, create new opportunities and shape change. With a growing need for skilled data scientists, the University of Notre Dame, in collaboration with AT&T, has announced its [new online master of science degree with a specialization in data science](#). Offered by the [Department of Applied and Computational Mathematics and Statistics](#), with the collaboration of the [Department of Computer Science and Engineering](#), the [Mendoza College of Business](#) and the [Department of Psychology](#), this degree program will prepare graduates for careers as data scientists in a wide range of industry fields including management, marketing, information technology, government policy, health care, finance, education and scientific research.



How Verizon Is Building A Big Data And AI Culture



CIO Network

Insights and ideas for technology leaders.

FULL BIO ✓

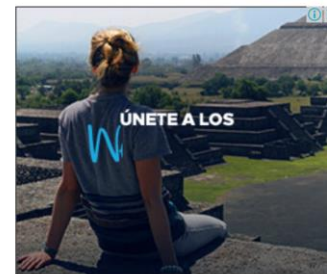
Opinions expressed by Forbes Contributors are their own.



CIO Central Guest, Contributor



The Verizon Center in Washington, Wednesday, Aug. 9, 2017. The [\[+\]](#)



Telecommunications has long been one of the most

Ciencia de Datos en el Mercado de Servicios - Telecomunicaciones



verizon✓

- Data Science and Cognitive Intelligence: IT, Marketing, Operaciones
- Global Supply Chain and Analytics Group: Procesos Internos, Delivery
- Big Data and Artificial Intelligence Systems: Startupera (Palo Alto)

Ciencia de Datos en el Mercado de Servicios - Telecomunicaciones



■ Clientes:

- ▷ Churn Rate (Grado de abandono): Se crean modelos analíticos para identificar potencial abandono del servicio.
 - ▷ Precio, Calidad de Servicio, Servicio insuficiente, Servicio al cliente
 - ▷ Soluciones:
 - ▷ Precios Dinámicos (Uber, Airbnb)
 - ▷ Calidad de Servicio – Infraestructura
 - ▷ Perfilamiento de clientes
 - ▷ Entrenamiento y personalización al cliente.

■ Infraestructura:

- ▶ Persistencia de la Red: Garantizar cobertura y servicio con grados de confianza.
 - ▶ Mantenimiento predictivo: Predecir cuando va a haber una baja de servicio por:
 - ▶ Sobrecarga en la línea
 - ▶ Eventos masivos (Terremotos)
 - ▶ Desastre Natural

- ▷ Ruteo inteligente:
 - ▷ Basados en los eventos en la predicción, hacer ruteos inteligentes.
 - ▷ Dado que el evento A va a pasar, la probabilidad de que B pase cruce el límite permitido.

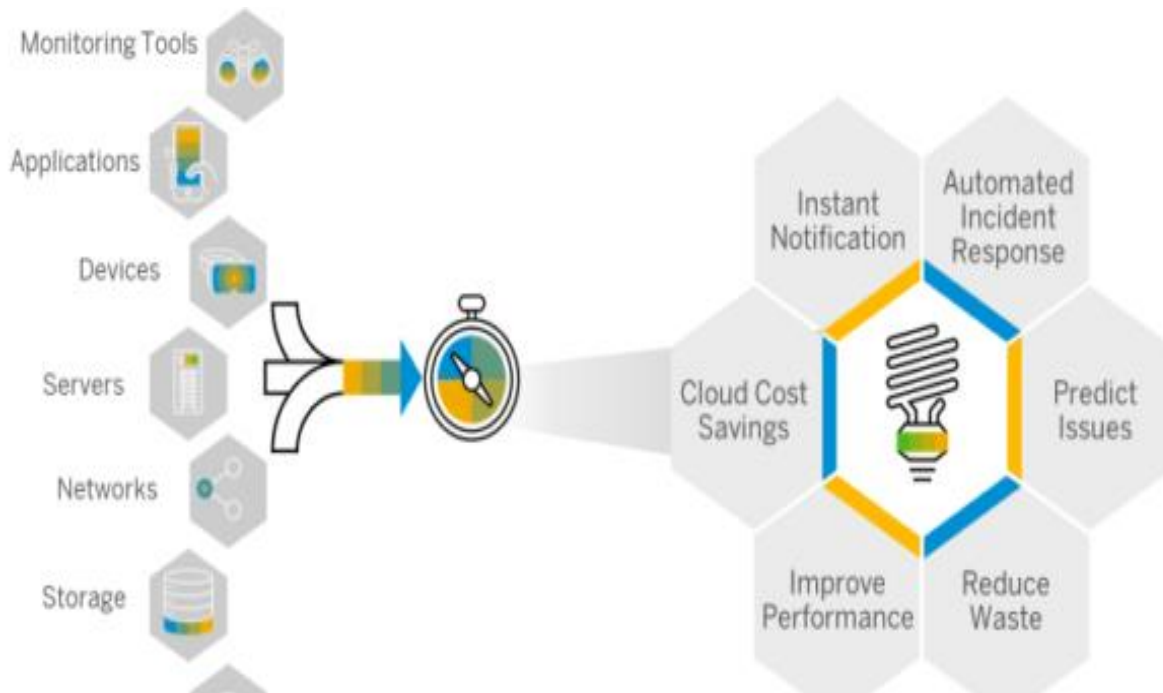
Ciencia de Datos en el Mercado de Servicios - Telecomunicaciones



■ Lluvia de ideas:

- ▷ Que otros problemas tienen las telecomunicaciones?
- ▷ Que tipo de servicio creen que le podrían proporcionar a Telmex o ATT?
 - ▷ Que lo hace único del mercado mexicano?
- ▷ Como cuantificarían el valor de ese servicio?

Ciencia de Datos en el Mercado de Servicios -IT



Ciencia de Datos en el Mercado de Servicios -TI



- Cuanto creen que le cueste a Amazon que un servidor no funcione
- Si el SAT tiene problemas con la red en Marzo, que creen que pase?
- ITOA es sorprendentemente difícil de implementar:
 - ▶ Analíticos + TI tradicional = Desastre!

Ciencia de Datos en el Mercado de Servicios - Entretenimiento

Giving Viewers What They Want



David Carr

THE MEDIA EQUATION

FEB. 24, 2013



Ciencia de Datos en el Mercado de Servicios - Entretenimiento

How Pop Artist Shelita Burke Uses Data Science And Blockchain To Her Advantage



Emily Blake, CONTRIBUTOR

I use data to tell stories about music. [FULL BIO](#) ✓

Opinions expressed by Forbes Contributors are their own.



Courtesy of Shelita Burke
Shelita Burke is an independent pop singer and data scientist in Los Angeles.

While some artists view their careers with a sort of "If you build it, they will come" mentality, pop singer-songwriter Shelita Burke definitely doesn't. As a data scientist and cryptologist, she has a

Ciencia de Datos en el Mercado de Servicios - Entretenimiento



Netflix:

- ▷ Filtros Colaborativos
- ▷ Motores de Recomendación
- ▷ Perfilamiento de clientes.

Hollywood/Música:

- ▷ Perfilamiento de clientes
- ▷ Análisis de sentimientos
- ▷ Regresiones de ganancia

Ciencia de Datos en el Mercado de Servicios - Entretenimiento



- Como le venderían un proyecto de Data Science a Televisa?
- Como le venderían un proyecto de Data Science a la LigaMx
- Cuales son los elementos de un hit?
 - ▷ Avengers
 - ▷ Star Wars
 - ▷ Pixar

The Future Of The Transport Industry - IoT, Big Data, AI And Autonomous Vehicles



Bernard Marr, CONTRIBUTOR
[FULL BIO](#) ✓

Opinions expressed by Forbes Contributors are their own.

Tremendous strides by innovators have resulted in some fascinating and intriguing enhancements to the transportation industry. In fact, in many ways, it feels like the future of the transportation industry is already here. These developments will cause transportation companies to rethink job descriptions and figure out when humans should be at the helm or when they should defer to smart machines for safety, cost savings and capabilities.



Ciencia de Datos en el Mercado de Servicios - Transporte



■ Rutas automatizadas:

- ▷ Perfilamiento de clientes
- ▷ Análisis de redes de comunicación.

Ciencia de Datos en el Mercado de Servicios - Transporte



Tráfico Dinámico

- ▶ Waze, Google Maps, etc
- ▶ “Event based traffic”
 - ▶ Basados en el historial de tráfico se puede predecir cuando se va a hacer un embotellamiento
 - ▶ Basados en el historial, se pueden planear los eventos y las nuevas construcciones.

São Paulo city government trials Big Data to improve public transport

A project in partnership with the World Bank wants to take bus management at the Brazilian capital to the next level



By [Angelica Mari](#) for [Brazil Tech](#) | February 20, 2015 -- 16:28 GMT (08:28 PST) | Topic: [Internet of Things](#)

Recommended Content:

White Papers: IBM - Developing a data integration and lifecycle management strategy for a hybrid environment

Download this IBM report to start addressing these specific questions: Where is the data that my organization needs? How will the data be assembled to create the information my organization needs?

[Join Today](#)



The city government of São Paulo is carrying out trials to improve the management of the local bus fleet of 15,000 vehicles by using data analytics.

Financed by the World Bank, the project is using technology supplied by Google-backed US startup Urban Engines, which provides both operators and officials of São Paulo's transport authority SPTrans with a comprehensive view of the buses and their ridership on a real-time basis.

Bus management in São Paulo is a resource-intensive exercise: to find out whether buses are on time, as well as the number of passengers riding on each bus, is a task that is often

RECOMMENDED FOR YOU

Simplify network management with FindIT Manager.

[White Papers provided by Cisco](#)

[READ MORE](#)

RELATED STORIES



CXO
LG mobile reshuffle points towards IoT focus

[Amazon](#)

Ciencia de Datos en el Mercado de Servicios - Transporte



- Cuales son los principales obstáculos para usar Data Science en la CDMX?
- Como lo resolverían?
- Vale la pena, o empezamos desde cero?

Ciencia de Datos en el mercado manufacturero - Automotriz



In 1913, the Ford Motor Company was at the forefront of car manufacture. Designing the reasonably-priced Model T to appeal to the masses and employing division of labour & mobilised assembly lines in the factory made Ford the largest automobile factory in the world at that time.

In 2007, the Ford Motor Company was in trouble. The end of 2006 financial year brought with it reports of a \$12.6 billion loss, the largest in the company's history. Yet, once again, forward-thinking innovation-this time, in the form of data analytics-led Ford back to the path of prosperity. By 2009, Ford was posting profits for the first time in 4 years, as well as launching 25 new vehicle lines. The same year they sold 2.3 million cars, the only company to exceed the 2 million mark since 2007. Last year, they won the 2013 INFORMS Prize for Company-Wide Efforts in Analytics & Data Science. So how exactly did Ford harness data analytics to bring them back from the brink and thrust them once again to the forefront of their industry?

BIG DATA, SUPPLY & DEMAND



SIGN UP FOR OUR NEWSLETTER

☒ Dataconomy ☒ Dataconomy Jobs ☒ Data Natives

Enter your email

SUBSCRIBE



POPULAR POSTS

WEEK

MONTH

ALL TIME



HOW BLOCKCHAIN CAN REVOLUTIONIZE
HEALTHCARE & MEDICAL RECORDS

BY DAAN AWEDD

Ciencia de Datos en el mercado manufacturero - Automotriz



■ - Qué vas a construir?:

- ▶ Ford tiene una de las líneas de ensamblado mas complejas del mundo.
- ▶ Que pasa si tienes control sobre lo que cada persona quiere en un auto.
- ▶ Utilizaron Facebook para minar las opiniones de las personas y diseñar así los autos.

Ciencia de Datos en el mercado manufacturero - Automotriz



■ Cómo lo vas a construir?

- ▷ Mantenimiento predictivo*
- ▷ Ford creó algoritmos que basados en las piezas en piso, pueden crear nuevos prototipos basados en los gustos del cliente.
- ▷ Como vas a entrenar a tus obreros en línea, y es ese entrenamiento útil?

Ciencia de Datos en el mercado manufacturero



■ - Manejo dinámico de inventarios:

- ▷ Modelos predictivos de demanda y uso de inventario.
 - ▷ Depende del producto, puede ajustarse dinámicamente
 - ▷ Predicción de consumo en punto de venta
- ▷ Predicción de demanda en cuestión de transporte

Ciencia de Datos en el mercado manufacturero

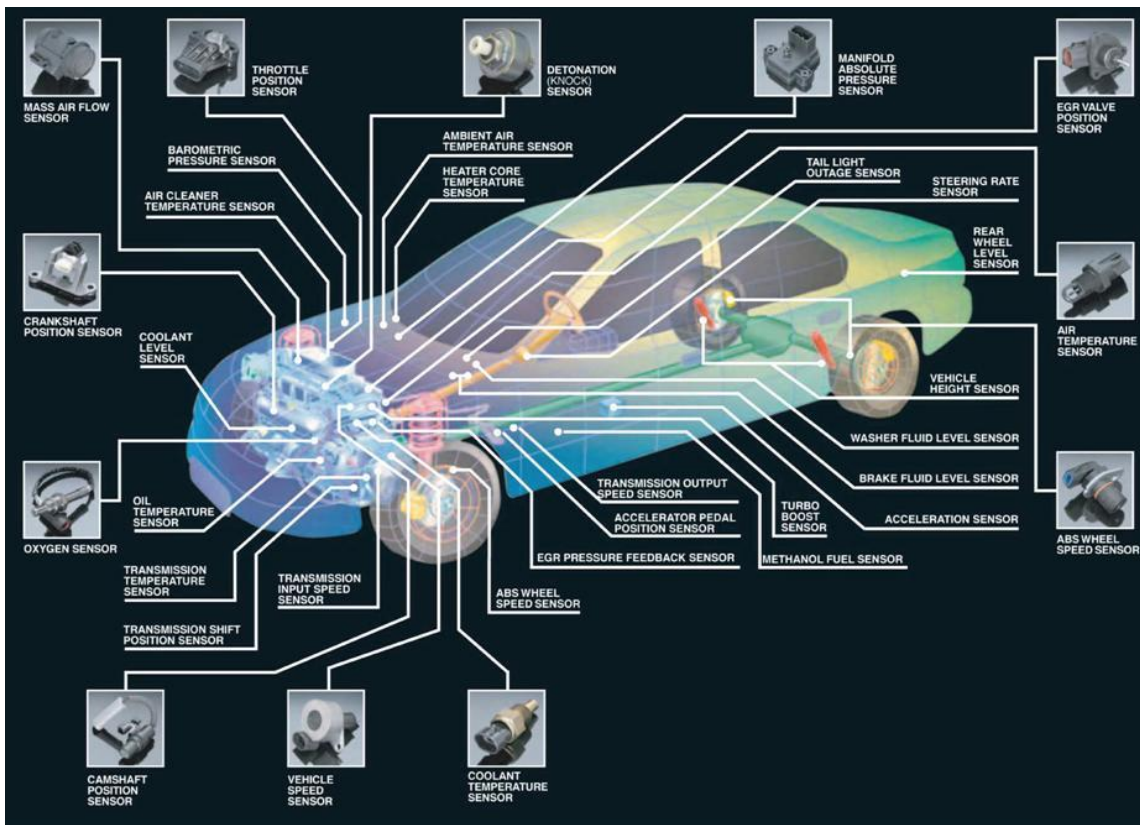


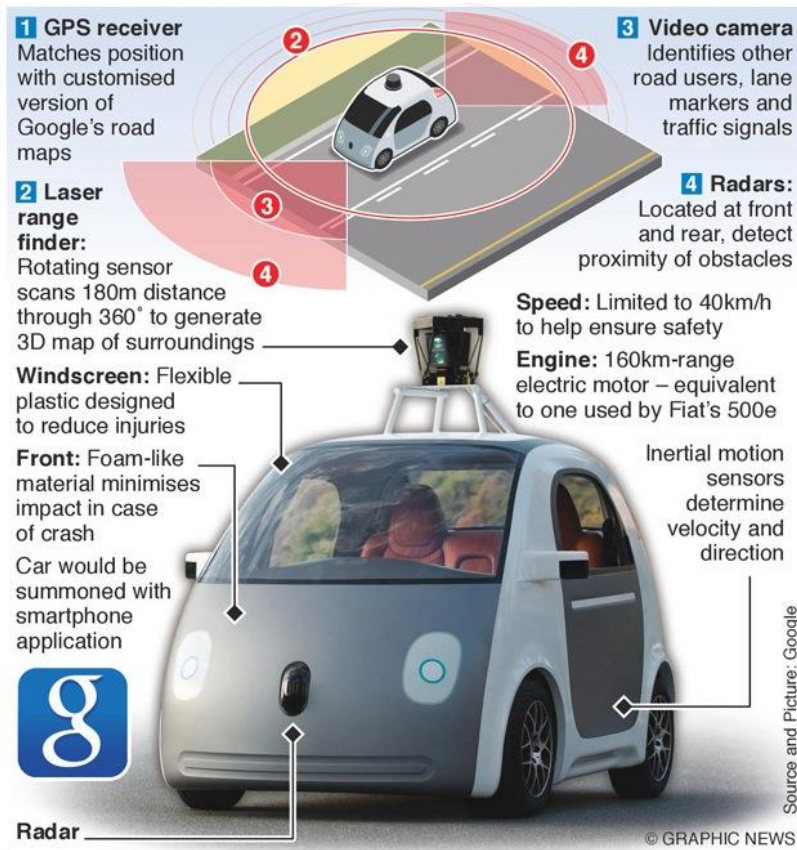
- ▷ Preguntas:
 - ▷ Que productos les parece que tendrán mayor demanda en estas tres fechas:
 - ▷ Navidad
 - ▷ Mundial de Futbol
 - ▷ Dia de acción de gracias.
 - ▷ Cómo medirían los efectos? Que necesitarían?

Internet de las cosas (IoT)

- No es propiamente Data Science.
- Pero sin DS, IoT no sirve absolutamente para nada.
 - ▷ Hay multiples empresas que ofrecen IoT as a service..... Y?
 - ▷ Un sensor, sin un análisis es inútil:
 - ▷ Tradicionalmente el análisis lo hacemos nosotros.







BIZ & IT—

How Amazon Go (probably) makes “just walk out” groceries a reality

Amazon's new age grocery likely wasn't possible even five years ago.

HAOMIAO HUANG - 4/10/2017, 6:30 AM



Enlarge / "Let's go shopping!" "No, let's Amazon Go shopping." "Dave, I hate your puns."

These days, most announcements by tech companies are pretty meh. Details either leak months ahead of time or reveal themselves to be pretty unimpressive. But lately, we've had some real surprises. Months ahead of releasing the Switch this spring, Nintendo decided the future of

Internet de las cosas (IoT)



Económico

- ▶ De que te sirven sensores de decenas de miles de dólares en un auto que cesta miles de dólares?
- ▶ El sensor debe ofrecer valor agregado que justifique su precio:
 - ▶ Pagarían ustedes 20,000 pesos más por una cámara de reversa
 - ▶ Pagarían ustedes 4,000 pesos más por una cámara de refrigerador.
 - ▶ Pagarían 300 pesos mas si no tienen que hacer fila en el super?

Internet de las cosas (IoT)

Bien diseñado:

- ▷ Si tienen sensores de hornos, vale la pena monitorearlo cada segundo?
 - ▷ Cada cuando? Como lo diseñarían?
- ▷ Vale la pena monitorear cada cliente que entra a la tienda?
 - ▷ Se debe?
- ▷ Vale la pena monitorear el rendimiento de los estudiantes por ciclo o por día?
 - ▷ Tablet, sentimientos, etc



UNIVERSIDAD
PANAMERICANA

Want big impact?
USE BIG IMAGE